氏　　名	漢　野　救　泰
生　年　月　日	
本　　籍	石川県
学 位 の 種 類	博士（工学）
学 位 記 番 号	博甲第529号
学位授与の日付	2002年9月30日
学位授与の要件	課程博士（学位規則第4条第1項）
学位授与の題目	有声音検出に基づく騒音下音声認識と確率モデルによる耐雑音性に関する研究
論文審査委員(主査)	船田哲男（工学部・教授）
論文審査委員(副査)	木村　春彦（工学部・教授）堀田　英輔（工学部・講師） 中山　謙二（工学部・教授） 下平　　博（北陸先端科学技術大学院大学・助教授）

学 位 論 文 要 旨

summary

The performance of speech recognition degrades remarkably due to the non-stationary noise or Lombard effect under heavy noisy environments. This is an existing important problem for the practical use of speech recognition system. In this work, firstly, we propose a feature parameter $Q_{L/W}$ to detect voiced sound periods using the prediction error by LPC(linear predictive coding) model in low-frequency band with the point aimed at the harmonic structure inherent in voiced sound. Secondly, we propose a method to recognize noisy Lombard speech, based on the detection of voiced sound periods and weighted distances for input speech. Then, the approach is applied to the fabric inspection system using speaker-dependent Lombard speech recognition. As the experimental results, this system achieves a high recognition performance and proves to be available for an improvement of operation efficiency in the inspection factory. Finally, we propose a method to modify the output probability at the state sensitive to noise by using weighted variance expansion based on the power of state or probability distribution, in order to make speech HMM robust to abrupt variations of noise. The effectiveness of this method is confirmed through the evaluation experiments of speaker-independent word recognition using noises of two factories.

1．研究の背景と目的

現状の音声認識技術では，静かな環境で発声すれば比較的高精度に音声認識ができるようになってきたが，騒音環境下での発声を認識するにはまだ不十分である．しかも，音声認識が有効と思われる環境のほとんどは雑音が大きく，その影響が無視できない．このため，近年の連続音声認識，自然発話認識へと研究の対象が進展する中にあっても，騒音環境下発声に関しては，限られた語彙数の離散単語においてさえ音声認識は難しく，特に非定常騒音下ではその認識率は極端に劣化する．このことは音声認識システムの実用化に向けて残された重要な課題となっている．

このような背景から，本研究は，騒音環境下での頑健な音声認識技術についての課題解決を目的としている．そして，工場での音声入力化の可能性を検討している．特に，手がふさがっている作業中に，機械の操作やコンピュータでのデータ入力などを作業者の音声で可能にすることは，効果的である．工場においては，騒音による非定常な雑音の混入でスペクトルが多様に変動して音声が不明瞭となる上，高騒音下での発声変形（ロンバード効果）により音声パターン自身が変動すると

いう課題が存在する．このため，本研究では，工場の騒音環境における非定常な雑音とロンバード効果に有効な音声検出手法，音声認識手法の開発を目指している．

2．騒音下音声認識における本研究の位置づけ

背景雑音の混入に対しては，入力音声の正規化，モデルの環境適応化に関する多くの研究が報告されている．定常雑音のように雑音条件が推定できれば各種正規化・適応化が有効であるが，実環境のほとんどは雑音が未知でかつ非定常なため対処が極めて難しい．特に，工場のような非定常騒音環境では，従来の手法だけでは改善効果が低い場合も多い．また，実際の高騒音環境下での発声による周囲の雑音が混入したロンバード音声に対しては，音声区間検出を含めた実用的に必要な認識評価がほとんどなされていない．

本研究では，ロンバード効果問題と雑音問題に対して，従来の対策手法ではまだ実用的に不十分な性能を改善する観点及びまだ確認されていない有効性を検証する観点から，雑音混入ロンバード音声に有効な有声音区間検出手法とこれに基づく特定話者用ロンバード音声認識手法を提供する．更に，未知の非定常雑音が混入した場合の音声認識に有効な単語 HMM 補償法を提案する．

3．有声音区間検出

まず，有声音特有の高調波構造に着目し，低域周波数帯における線形予測（LPC）モデルの予測残差を利用することによって，有声音区間を検出する手法（ピッチ対応型低域 LPC 分析手法）及び特徴パラメータ（低域 LPC 補正適合度 $Q_{L/W}$ ）を提案した．本手法では，スペクトル上でピッチ周波数とその高調波に対応するピークを，全極型モデルの極とみなして分析を行い，その適合の度合いから有声音を検出する．LPC 適合度は正規化残差パワーの逆数の対数として定義しており，低域 LPC 適合度 $Q_{L/L}$ に対して低域パワーと広帯域パワーとの比で補正したものが $Q_{L/W}$ である．そして，LPC 分析の対象とする周波数帯域を低域に限定することにより，広帯域を使用する場合と比較して有声音と雑音の分離性能を向上させるとともに，計算量低減も可能にした．

表 1　有声音フレームの検出結果

特徴パラメータ		検出率(%)
広帯域 LPC 適合度	$Q_{W/W}$	23.7
低域 LPC 適合度	$Q_{L/L}$	65.3
低域 LPC 補正適合度	$Q_{L/W}$	82.8
自己相関係数ピーク値	r_p	74.7
自己相関係数ピーク補正値	r_c	73.3

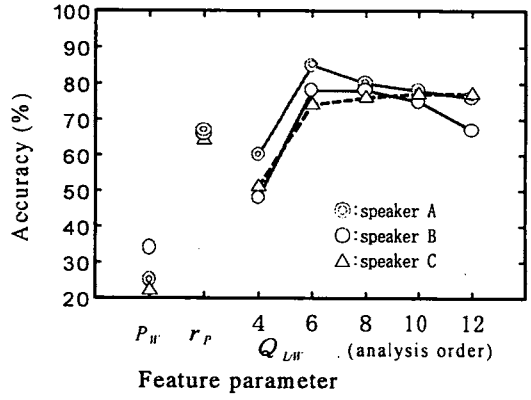

図 1　3 名の話者の有声単語区間検出実験の結果

本手法の有効性を，機械加工工場の非定常騒音環境下での発声単語を用いた有声音と雑音の分離性能評価実験，有声音フレームの検出実験（表 1），有声単語区間の検出実験（図 1）により評価した．高騒音下では無声音は検出困難なため，単語音声認識の前処理である音声区間検出としては，単語の最初の有声音区間の始端から最後の有声音区間の終端までを単語区間として検出するのが実用的と考え，この有声音に基づく単語区間を有声単語区間として定義している．

表 1 より，$Q_{L/W}$ は SIFT 法に基づく自己相関係数ピーク値 r_p ，及びその補正値 r_c 以上の検出率を示している．また，図 1 から分析次数 6 次以上でいずれの話者の有声単語区間検出率においても $Q_{L/W}$ が r_p を上回っている．そして，比較的ピッチ周波数の高かった話者 A は，話者 B，C よりも，

低次数（4 次，6 次）で効果が高かった．このことは，本ピッチ対応型低域 LPC モデルが正しく機能しており，想定される極数から得られる次数以上であれば，有声音に適合し雑音中からの有声単語区間検出が可能であることを示している．

評価実験の結果，$Q_{L/W}$は，1 パラメータで高調波構造の LPC 適合，パワー値が考慮されており，さまざまな非定常雑音に対して低く安定しているため，r_p よりも騒音下発声における有声音検出用特徴パラメータとして有効であることがわかった．

4．雑音混入ロンバード音声の認識

$Q_{L/W}$を用いた有声音区間検出と入力音声に対する距離重み付け DP マッチングに基づく実用に適した雑音混入ロンバード音声認識方式を提案するとともに，最も性能が期待できる方式として，標準パターンに実環境下発声音を使用できる認識システムを検討した(図 2)．

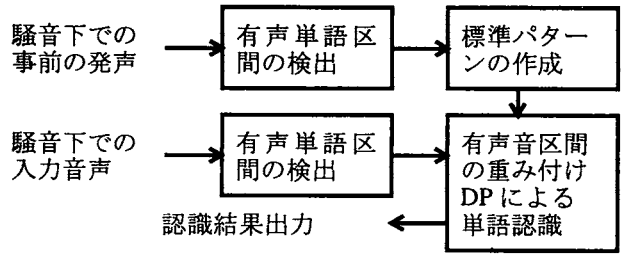

図 2　ロンバード音声認識方式

そして，工場で発声した単語の認識実験により，騒音下発声音を標準パターンとして利用する本認識手法の有効性を各種距離尺度（CEP，RPS，WGD）を用いて評価した．その結果（表 2）から，本研究提案の有声単語区間利用は従来の音声区間利用よりも各距離尺度で認識性能が高く，有声単語区間利用の有効性を確認した．また，$Q_{L/W}$による本有声単語区間検出手法が SIFT 法に基づく手法より優れていることが認識性能からも明らかになった．距離尺度としては，重み付け群遅延距離尺度 WGD が最も高い認識性能を示し，実用的には，標準パターンとして騒音環境下で衝撃音を含まないように発声音を有声単語区間で登録する方法が最も有効であった．また，入力音声の有声音区間について距離を重み付けすることで音声認識性能の向上が確認できた．

表 2　騒音下発声音の各種距離尺度・端点フリー範囲での単語認識率(%)
(模擬騒音下発声音を標準パターンとして用いた場合　Q:$Q_{L/W}$　S:SIFT 法)

(a)　有声単語区間を使用

始端 終端	固定 固定	±25ms ±25ms	±25ms ±50ms	±50ms ±50ms	±50ms ±100ms	±100ms ±100ms
CEP　Q	76.7	78.6	**79.4**	78.6	78.1	73.9
CEP　S	71.1	72.2	**74.7**	74.2	73.1	71.1
RPS　Q	83.3	83.9	84.4	**85.3**	83.3	79.4
RPS　S	77.2	78.3	79.2	80.3	**80.6**	76.9
WGD　Q	85.8	86.9	86.9	**87.8**	87.2	84.7
WGD　S	79.4	80.8	81.7	81.7	**83.9**	81.1

(b)　無声音を含む音声区間を使用

始端 終端	固定 固定	−50ms~0 0~50ms	−100ms~0 0~100ms	−150ms~0 0~150ms	−200ms~0 0~200ms
CEP　Q	65.6	74.2	76.9	**77.5**	76.4
CEP　S	60.6	69.2	72.5	**73.3**	73.1
RPS　Q	71.4	79.2	82.5	**83.3**	82.5
RPS　S	65.6	73.3	77.8	**80.3**	80.0
WGD　Q	71.6	80.3	85.3	**85.6**	84.7
WGD　S	66.7	74.7	81.1	**83.1**	82.8

更に，これらの評価結果を基に，特定話者ロンバード音声認識による音声入力検反システムを開発した．検査工場において，音声により織物欠点名を入力する動作実験を行った結果，必要とされる認識性能を上回る高認識率を達成し，検査工程での操作性向上が可能なことがわかった．

5．HMM の耐雑音性改善

工場のような実環境下では，非定常な高騒音の発生など周囲の状況の変化により，雑音のスペクトルや SN 比が急激に変動することがある．このような雑音条件変動に対する HMM の耐雑音性を向上させるため，各状態あるいは各分布のパワーによる重み付け分散拡大により，雑音の影響を受けやすい状態からの出力確率を制御する手法を提案した（図３）．

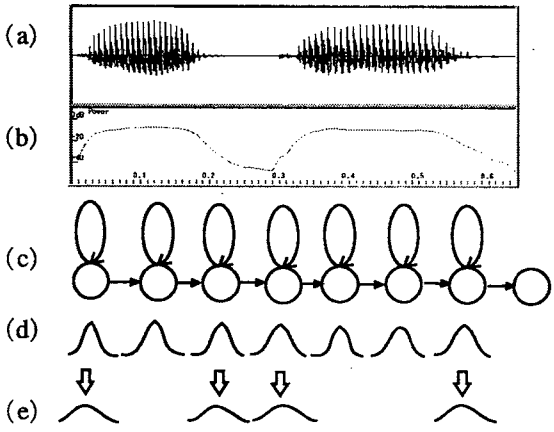

図３　パワーによる重み付け分散拡大の概念図
((a):音声波形‘東京’, (b):パワー, (c):単語 HMM, (d):出力確率分布, (e):低パワーの状態での分散拡大)

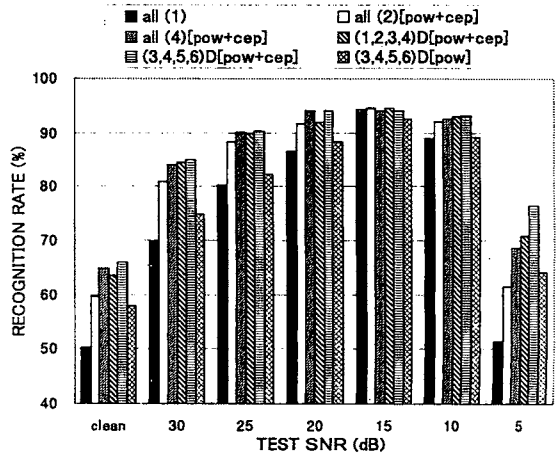

図４　15dB-HMM の学習時とは異なる雑音での拡大率一定と重み付けとの認識性能比較

学習時とは異なる雑音条件における本手法による認識性能改善効果を，語彙数 50 の単語音声と 2 種類の工場の雑音を使用した不特定話者単語認識実験で評価した．HMM としては，クリーン音声 HMM と 2 種類の雑音付加音声 HMM を使用した．15dB の SN 比の雑音付加音声で学習した HMM（15dB-HMM）と 5 種類の SN 比の雑音付加音声で学習した HMM（5SNR-HMM）である．そして，定常時が比較的静かな環境であればクリーン音声 HMM，騒音環境では雑音付加音声 HMM を使用し，その雑音条件での最高の認識性能を維持するとともに，異なる雑音条件となった場合でも従来の性能低下を改善する方法について検討した．

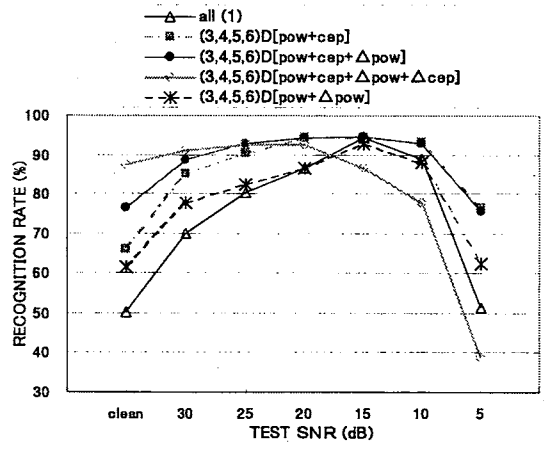

図５ 15dB-HMM の学習時とは異なる雑音での分散拡大の特徴パラメータと認識性能の関係

図４，５は 15dB-HMM を使用した実験で得られた単語認識率を示している．all(・)は拡大率一定を表し，(3,4,5,6) D はパワーの低い分布に対して拡大率を大きくしたことを表している．また，[・]は分散拡大したパラメータを表している．

評価実験の結果，いずれの HMM においても，以下の確認ができた．(1)本手法により，雑音条件の変動に対して広範囲の SN 比における平均認識率が向上し，音声 HMM の耐雑音性が改善できた．(2)特に，学習時より低い SN 比の雑音付加音声に対して，重み付け分散拡大は拡大率一定よりも認識性能を顕著に改善できることが明らかになった．以上のことは，モデルの尤度計算の観点からは，雑音付加・変動に対して信頼性の低い出力確率分布を有する状態からの出力確率のモデル

尤度への寄与を低減させる効果が現れていると考えられる．

また，分散拡大が有効なパラメータとしては，以下のことが確認できた．

●クリーン音声 HMM

パワー(pow)＋ケプストラム(cep)の分散拡大が，他のパラメータの分散拡大よりも SN 比の低下に対して有効である．

●雑音付加音声 HMM

認識時の雑音の種類が学習時と同じか異なるかで改善性能に違いが生じる．同種の雑音では，学習時の SN 比±5dB で pow＋△pow，それ以上に SN 比が異なる場合及び異種雑音では pow＋cep＋△pow の分散拡大による改善効果が高かった．

6．結論

本研究では，まずピッチ対応型低域 LPC 分析手法と低域 LPC 補正適合度が騒音下での有声音検出に有効であることを示した．そして，この手法に基づく有声単語区間を利用した有声音区間重み付け DP マッチング手法が実環境下での雑音混入ロンバード音声認識に効果的であることを確認した．更に，非定常騒音環境での雑音の急激な変動に対して，重み付け分散拡大手法は，単語音声 HMM の学習時や適応時とは異なる雑音条件における耐雑音性を改善できることを明らかにするとともに，改善に最適な分散拡大条件に関する知見が得られた．

学位論文審査結果の要旨

平成 14 年 7 月 29 日に第 1 回学位論文審査委員会を開催，7 月 30 日に口頭発表，その後に第 2 回審査委員会を開催し慎重審議の結果，以下の通り判定した。

高騒音環境下では，非定常な雑音の混入と発声変形（ロンバード効果）により，音声認識性能の低下が顕著であり，音声認識システムの実用化に向けてこの性能低下を軽減することが重要な課題となっている。本研究では，(1) 有声音特有の高調波構造に着目し，低域周波数帯における線形予測モデルの予測残差を利用する有声音検出法を提案し，雑音環境下におけるその有効性を確認した。(2) 入力音声に対する距離重み付けに基づく実用に適した雑音混入ロンバード音声認識手法を提案し，特定話者ロンバード音声認識による音声入力検反システムを開発した。検査工場において，音声による織物欠点名を入力する動作実験で高認識率を達成し，検査工程での操作性向上が可能なことを示した。(3) 雑音条件変動に対する音声確率モデル (HMM) の耐雑音性を向上させるため，各状態あるいは各分布のパワーによる重み付け分散拡大により，雑音の影響を受けやすい状態からの出力確率を制御する手法を提案し，雑音条件の変動に対して広範囲の SN 比における平均認識率を向上して，音声 HMM の耐雑音性を改善した。さらに，学習時より低い SN 比の雑音付加音声に対して，パワーによる重み付け分散拡大の方が拡大率を一定とするよりも認識性能を顕著に改善できることを示した。以上のように，本論文は音声認識における耐雑音性向上に関する有効な方法を提案しており，博士（工学）に値するものと判定した。